



# Motion detection and the coincidence of structure at high and low spatial frequencies

Peter J. Bex<sup>\*</sup>, Steven C. Dakin

*Institute of Ophthalmology, 11-43 Bath Street, London EC1V 9EL, UK*

Received 31 January 2002

---

## Abstract

We used filtered random dot kinematograms and natural images to examine how motion detection depends the relative locations of structures defined at low and high spatial frequencies. The upper displacement limit of motion ( $D_{\max}$ ), the lower displacement limit ( $D_{\min}$ ) and motion coherence thresholds were unaffected by the degree of spatial coincidence between high and low spatial frequency structures i.e. whether they were consistent or inconsistent with a single feature. However motion detection was possible between band-pass filtered random dot patterns whose peak frequencies were separated by up to 4 octaves. The first result implicates spatial frequency selective motion detectors that operate independently. The second result implicates a motion system that can integrate the displacements of edges defined by widely separated spatial frequencies. Both are required to account for the two results, and they appear to operate under very similar conditions.

© 2003 Elsevier Science Ltd. All rights reserved.

*Keywords:* Motion; Random dot kinematogram;  $D_{\max}$ ;  $D_{\min}$ ; Spatial frequency; Edges; Features

---

## 1. Introduction

An interesting property of the natural images that the visual system has (presumably) evolved to process (Attneave, 1954; Barlow, 1961) is that spatial structure is often correlated across spatial scales, so that the apparent location of contours in images from different spatial frequency bands are often coincident. In the present study we ask whether this correlation is important for the detection of motion.

Standard computational approaches to biological motion detection are based on autocorrelation (Reichardt, 1961; van Santen & Sperling, 1985), motion energy (Adelson & Bergen, 1985; Watson & Ahumada, 1985) or spatio-temporal gradient (Johnston, McOwan, & Benton, 1999; Johnston, McOwan, & Buxton, 1992) techniques. Each technique provides an independent *local* estimate of the direction of motion, i.e. within a small area of the visual field, and often over a limited

range of spatial and temporal frequencies (Adelson & Bergen, 1985; Johnston et al., 1999; Watson & Ahumada, 1985). To derive an estimate of the *global* pattern of movement, the output of such local direction estimates must be combined to provide an overall estimate of motion across a large area of visual space, and over a range of spatio-temporal frequencies in broad-band images. It is not known how this class of motion detector is affected by any correlation of spatial structure across scales in natural images. Other models of motion detection code the change in position of spatial primitives such as zero crossings (Marr & Hildreth, 1980; Ullman, 1979), zero bounded regions (Watt & Morgan, 1985), luminance peaks (Eagle & Rogers, 1996) or points of phase alignment (Morrone & Burr, 1988). For these models, it is the correlation of structure across spatial frequencies that define the spatial primitives on which motion detectors operate.

The perception of motion has been extensively studied with random dot kinematograms (RDKs). In these stimuli, spatially shifted versions of random dot images presented in rapid succession give rise to the appearance of smooth motion up to a maximum displacement size known as  $D_{\max}$  (Braddick, 1974). Many studies have examined how  $D_{\max}$  varies as a function of the

---

<sup>\*</sup> Corresponding author. Tel.: +44-207-608-4015; fax: +44-207-608-6983.

E-mail address: [p.bex@ucl.ac.uk](mailto:p.bex@ucl.ac.uk) (P.J. Bex).

spatial frequency content of filtered RDKs (Bex, Brady, Fredericksen, & Hess, 1995; Bischof & Di Lollo, 1990, 1991; Brady, Bex, & Fredericksen, 1997; Chang & Julesz, 1983a,b; Cleary & Braddick, 1990a,b; Eagle, 1998; Eagle & Rogers, 1996; Hess, Bex, Fredericksen, & Brady, 1998; Ledgeway, 1996; Morgan & Mather, 1994), the element size (Morgan, 1992; Morgan & Fahle, 1992), or both (Morgan, Perry, & Fahle, 1997; Smith & Ledgeway, 2001).

In band-pass filtered RDKs,  $D_{\max}$  broadly obeys the half cycle limit (in which the directional response of a narrow-band motion detector reverses when a periodic stimulus is displaced by more than one half cycle) allowing for the contribution of components at orientations non-orthogonal the axis of displacement which are displaced by less than one quarter cycle within  $\pm 90^\circ$  of this axis (Bischof & Di Lollo, 1991).  $D_{\max}$  is unaffected by small increases in element size (resulting from either direct size increases or moderate low-pass spatial frequency filtering), but steadily increases as the element size exceeds a critical point, typically beyond sizes of about  $10'$  (Morgan, 1992) or low-pass filter cut-offs below 3–4 c/deg (Cleary & Braddick, 1990b). These results are consistent with motion models that are based on feature correspondences (Eagle & Rogers, 1996), possibly following initial low-pass filtering by the visual system (Morgan, 1992; Morgan & Mather, 1994). The results are also consistent with the responses of a population of motion energy detectors tuned to a narrow range of frequencies in which noisy, high contrast directional signals at high spatial frequencies mask coherent directional signals at lower spatial frequencies (Bex et al., 1995; Cleary & Braddick, 1990b; Hess et al., 1998).

The observation that  $D_{\max}$  is large for high-pass filtered RDKs composed of large elements cannot easily be explained by the pooled response of a population of motion energy detectors and is best explained by models based on the correlation of edges (Eagle & Rogers, 1996; Morgan et al., 1997; Smith & Ledgeway, 2001). Recent evidence showing that inverting the contrast polarity of filtered RDKs produces reversals in the apparent direction of motion at small element sizes, but not large elements sizes (Smith & Ledgeway, 2001), suggests that observers utilise both motion energy and feature-based direction discrimination strategies under viewing conditions that overlap much more than was previously supposed (Boulton & Baker, 1994, 1993; Baker & Hess, 1998; Bex & Baker, 1997; Lu & Sperling, 1995).

### 1.1. Rationale

An important distinction between energy and feature-based approaches to motion detection is that feature-based models extract the location of edge tokens *before* encoding their change of position over time, whereas

spatio-temporal filters code direction *independently* of any edge coding processes and possibly independently of other detectors operating at different spatio-temporal frequencies. We used this distinction to determine the conditions under which energy or edge based motion detection strategies support motion detection in broad-band random dot images and natural scenes. We devised hybrid stimuli that were the sum of a low and a high spatial frequency band-pass filtered image pair, the rationale is illustrated graphically in Fig. 1.

The panels show one dimensional luminance profiles of noise images or band-pass filtered noise and the locations of zero crossings (Marr & Hildreth, 1980). Panels A and B show a random binary noise image, Noise X (Ai and Bi), followed by band-pass filtered images centred at high (48 cycles per image, cpi, Aii) or low (8 cpi, Bii) spatial frequencies, followed by the locations of the signed zero crossings in the filtered images (Aiii and Biii). In panel C, an independent random binary noise image, Noise Y (Ci) is band-pass filtered (48 cpi, Cii). Panels Di and Ei show images that are the sum of low and high spatial frequency band-pass filtered images, together with the zero crossings in the summed image (Dii and Eii). In panel D, the high and low spatial frequency band-pass images are both derived from the same source image and the summed image (Di) closely resembles the source image (Ai or Bi). Furthermore the locations of the zero crossings in all images whether filtered, unfiltered or compound sum are identical (Aiii, Biii and Dii). However, in panel E, the band-pass images are derived from independent source images (low SFs from Noise X, high SFs from Noise Y). In this compound image, the zero crossings (Eii) are more numerous and are not coincident with those in either source image.

Experiments 1 and 3 were based on the fact that spatio-temporal motion energy filters operate independently on a narrow range of spatial frequencies (such as Aii, Bii, or Cii), so any correlation across scales is unimportant. Motion detection along these principles should be unaffected by any coincidence of structure across spatial scales. However, the more numerous edges in compound images with uncorrelated structure at high and low spatial frequencies (Eii) causes feature-matching operations to collapse at shorter values of  $D_{\max}$  owing to the increased probability of false correspondences between any of the more numerous like-signed edges. To contrast these approaches to motion processing, we measured direction discrimination with compound images that were the sum of low and high spatial frequencies from either the same or independent images.

Given that spatio-temporal motion energy filters operate on a band-pass image representation, they should respond weakly if at all to images from different spatial frequency bands (such as Aii and Bii). However, the

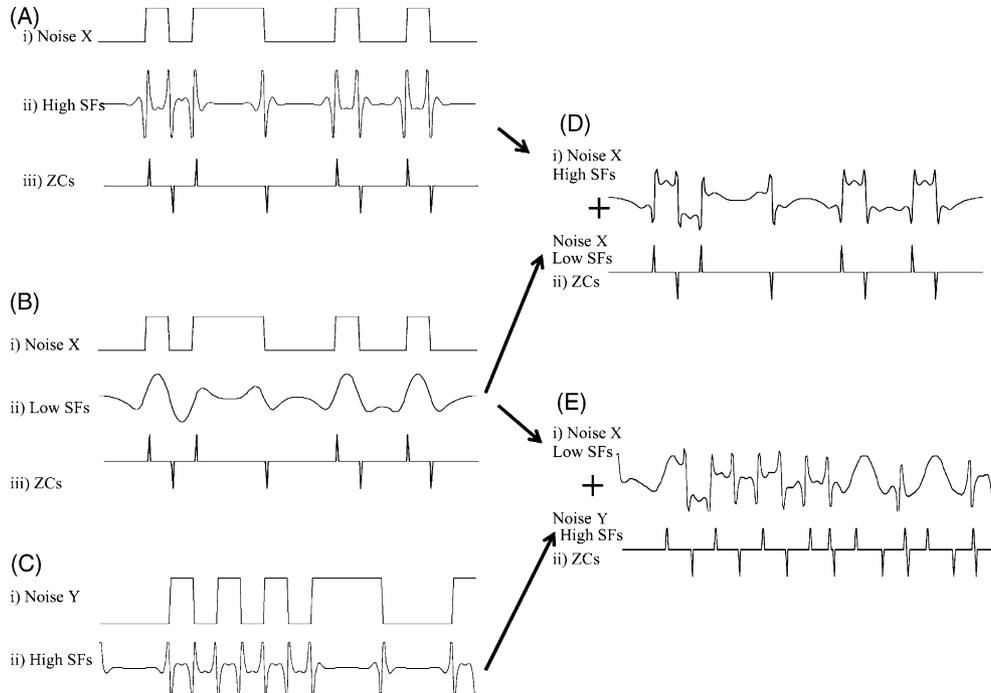


Fig. 1. Rationale. One dimensional luminance profiles and edge locations in filtered and unfiltered noise images. (Ai) 1D luminance profile of binary noise X, followed by (Aii) a band-pass filtered version of it (fpeak @ 48 cpi) then by (Aiii) the zero crossings in (Aii). B is the same as A, except that the peak spatial frequency of the band-pass filtered image in (Bii) is 8 cpi. (Ci) A different noise source Y is filtered (fpeak @ 48 cpi) to produce (Cii). (Di) shows the luminance profile produced by summing the high and low spatial frequency components of noise X (Aii and Bii respectively). (Dii) shows the zero crossings in (Di). (Ei) shows the luminance profile produced by summing the low spatial frequency structure from noise X with the high spatial frequency structure from noise Y (Cii and Bii respectively). (Eii) shows the zero crossings in (Ei). The compound wave in (Di) resembles its source noise X (Ai and Bi), and the zero crossings occur at the same locations in the source images (Aiii), (Biii) and the summed image (Dii). The zero crossings in (Eii) are more numerous and do not correspond to those in the source images. See text for details.

correspondence between the edges in some band-pass filtered images (Aiii and Biii) should support motion detection based on coding the locations of spatial primitives even when there is no spatial frequency structure common to both images in a RDK. Therefore in Experiment 2, we measured direction discrimination between band-pass filtered images as a function of the spatial frequency overlap between the two frames of a RDK.

## 2. Experiment 1: Image structure at high and low spatial frequencies

### 2.1. Methods

#### 2.1.1. Apparatus

Stimuli were generated on a Macintosh G4 computer with software adapted from the VideoToolbox routines (Pelli, 1997) and were displayed on a LaCie Electron-Blue 22 in. monitor in greyscale at a frame rate of 75 Hz. Stimuli had a mean luminance of 50 cd/m<sup>2</sup> and were presented at 75% Michelson contrast. The luminance of the display was linearized with pseudo-12 bit resolution (Pelli & Zhang, 1991) in monochrome and calibrated with a Minolta photometer. Images were presented in

greyscale by amplifying and sending the same 12-bit monochrome signal to all RGB guns of the display. The display subtended 26° horizontally (832 pixels), 19.5° vertically (624 pixels), and was 86 cm from the observer, in a dark room.

#### 2.1.2. Stimuli and procedure

Stimuli were presented in a central circular window subtending 8° (256 pixels), the edges of which were smoothed with a raised cosine subtending 0.5° (16 pixels). RDKs were presented for 213 ms (16 video frames at 75 Hz) with an abrupt spatial displacement at 106 ms. The onset and offset of the RDK was smoothed with a raised cosine temporal envelope lasting 40 ms. A central fixation cross was present before and after each trial, but not while the RDK was presented.

The natural images were drawn at random from a database of calibrated natural images (van Hateren & van der Schaaf, 1998) that are freely available to download from this web site: <http://hlab.phys.rug.nl/archive.html>. The source images were imported as 16 bit numbers corresponding to a rectangular image of size 1536 × 1024 pixels, then down-sampled to 8 bits following our filtering. The angular resolution of each image pixel was approximately 2' and this resolution was

maintained in our experiments. On every trial an image was selected at random and a square region of  $256 \times 256$  pixels was cropped from a random region within it, along with an adjacent area to allow for spatial displacement.

The RDKs were binary noise fields in which each element was either black or white, with equal probability. There were five element sizes ( $0.0625^\circ$ ,  $0.125^\circ$ ,  $0.25^\circ$ ,  $0.5^\circ$  and  $1^\circ$ ). A new random noise sample was generated every trial. The images were then digitally filtered using conventional techniques (Press, Teukolsky, Vetterling, & Flannery, 1992) with logarithmic exponential filters, which have the advantage of shorter tails than Laplacian or Gaussian filters:

$$A(f) = \exp\left(-\frac{|\ln(f/F_{\text{peak}})|^3 \ln 2}{(b_{0.5} \ln 2)^3}\right) \quad (1)$$

where  $F_{\text{peak}}$  specifies the peak frequency and  $b_{0.5}$  the half bandwidth of the filter in octaves. There were two centre frequencies, low ( $F_{\text{peak}} = 1$  c/deg) and high ( $F_{\text{peak}} = 6$  c/deg), the half with half height bandwidth in both cases was 1 octave.

There were four experimental conditions:

1. *Low SF*: low spatial frequencies only.
2. *High SF*: high spatial frequencies only.

3. *Broad SF coherent*: low and high frequencies from the same noise sample summed to generate a single image.
4. *Broad SF incoherent*: low and high frequencies from independent random noise samples summed to generate a single image.

Illustrations of representative stimuli are shown in Fig. 2.

A two-alternative forced choice direction discrimination task was used to determine  $D_{\text{max}}$ . Observers indicated whether the image had shifted to the left or right. Auditory feedback was provided following incorrect responses. The magnitude of the spatial displacement was controlled by a QUEST staircase (Watson & Pelli, 1983). Each staircase was initialised with a random starting level and concentrated observations around a displacement producing 75% correct responses. There were 32 trials for every condition of element size combined with the four filter combinations. All conditions were interleaved in a single run and each run was completed a minimum of 4 times. A Cumulative Normal function was fitted to the combined data of the runs by least  $\chi^2$ , weighted by the binomial standard deviation of each data point.  $D_{\text{max}}$  was inferred as the 75% correct point of this function, together with 95%

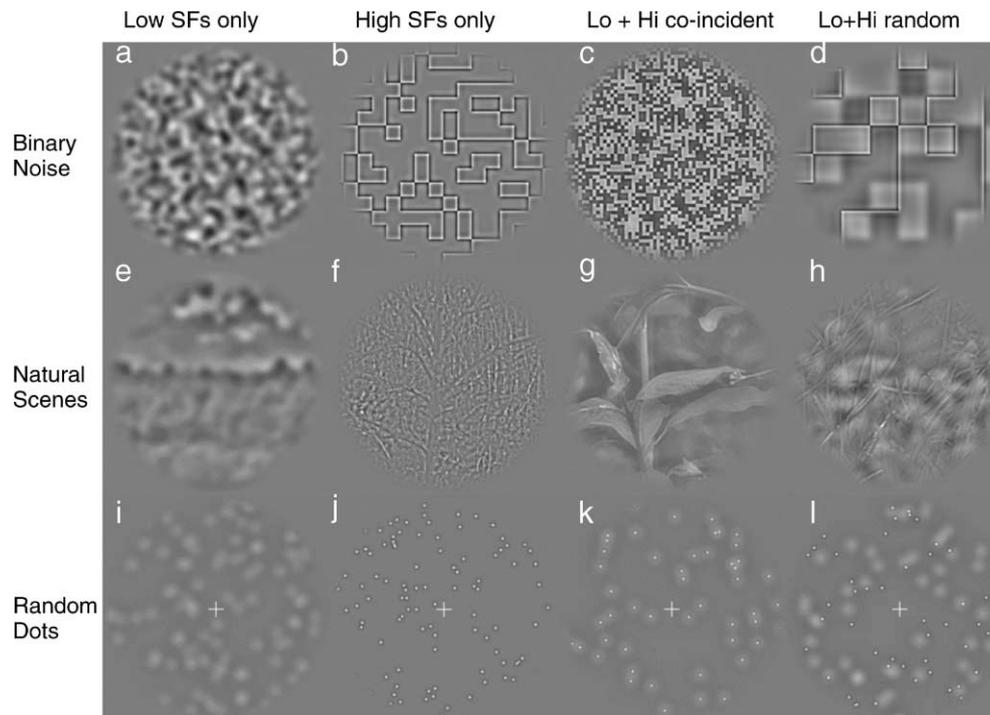


Fig. 2. Examples of the stimuli. The top row (a–e) shows filtered binary noise images, the middle row (e–h) shows filtered natural scenes and the lower row (i–l) shows filtered random dot images. The first column shows images containing only low spatial frequencies ( $F_{\text{peak}} = 1$  c/deg under our viewing conditions). The second column shows images containing only high spatial frequencies ( $F_{\text{peak}} = 6$  c/deg). The third column shows images containing low and high spatial frequencies in which the edges are coincident (i.e. the sum of low and high spatial frequency components of the same image). The fourth column shows images containing low and high spatial frequencies in which the edges are random (i.e. the sum of low and high spatial frequency filtered components of different images). All source images are  $256 \times 256$  pixels. The elements sizes of the binary noise images are: (a) 2, (b) 16, (c) 4, (d) 32 pixels, where 1 pixel was  $1.875'$  under our viewing conditions. See text for details.

confidence intervals by standard methods (Press et al., 1992).

In order to equate the visibility of high and low spatial frequency structure in our stimuli, we first collected contrast detection thresholds for each observer. Stimuli and procedure were as described above, except that QUEST controlled the contrast of the natural images or the filtered noise images and the observers' task was to indicate whether the stimulus appeared on the left or right of fixation, with feedback for incorrect responses. The relative contrast thresholds for low and high spatial frequencies (0.34:0.66, for PB and 0.33:0.67 for RW, respectively) were used to scale the relative contrasts of these images in the RDK, whose final Michelson contrast was fixed at 75%.

## 2.2. Results

Fig. 3 shows  $D_{\max}$  for natural images for two observers for four conditions shown on the  $x$ -axis. Repeated measures ANOVA showed a significant difference between conditions ( $F_{(3,3)} = 21.4$ ,  $p < 0.025$ ).  $D_{\max}$  is significantly lower for high SF only noise ( $p < 0.01$ , by linear contrasts) and not significantly different among low SF, broad SF coherent and broad SF incoherent noise. Thus there is no effect of the alignment of edges at low and high spatial scales. The value of  $D_{\max}$  is considerably larger than the values recorded with noise images below, even for high SF natural images, but is in line with values of  $D_{\max}$  reported with filtered  $1/f$  noise images (Bex et al., 1995; Hess et al., 1998) which have a similar amplitude spectrum to the images used in the present study (Burton & Moorhead, 1987; Field, 1987).

Fig. 4 shows  $D_{\max}$  as a function of element size for two observers. The four conditions are shown by different symbols: open squares show high SF stimuli only;

open circles show low SF only; filled squares show high and low SFs from the same noise sample; filled circles show high and low SFs from independent noise samples.  $D_{\max}$  is lowest for high SF only noise and approximately equal for low SF, broad SF coherent and broad SF incoherent noise. There is no effect of the alignment of edges at low and high spatial scales at any element size.

Our values of  $D_{\max}$  are slightly higher than some previous estimates of  $D_{\max}$  under similar conditions (Morgan, 1992), but are comparable to others (Morgan et al., 1997). The latter paper discusses potential causes of the discrepancy and we also note that our larger image size ( $8^\circ$  compared to  $5^\circ$ ) could also have contributed to the difference through probability summation of directional responses. The results confirm previous studies showing that  $D_{\max}$  is initially unaffected by increases in element size, then rises steadily above a critical size that depends on either the initial filtering by the visual system (Morgan, 1992) or of the stimulus. The knee point in the functions at element sizes of around  $10\text{--}20'$ , is in close agreement with all previous studies (Morgan, 1992; Morgan et al., 1997; Smith & Ledgeway, 2001).

To estimate the relative number of edges/zero crossings in hybrid images from identical or independent noise sources, we ran 250 simulations with one dimensional images at an element size of 16 pixels in a 256 pixel strip, and filters at 8 and 48 cpi (equivalent to 1 and 6 c/deg under our viewing conditions). A threshold value determined the minimum intensity difference between adjacent pixels that could be classed as an edge. Hybrid images from independent noise sources at low and high SFs had between 2 (threshold = 0) and 1.4 (threshold = 1 standard deviation of the image intensity) times as many edges as hybrid images from the same noise source, depending on the threshold value. However,

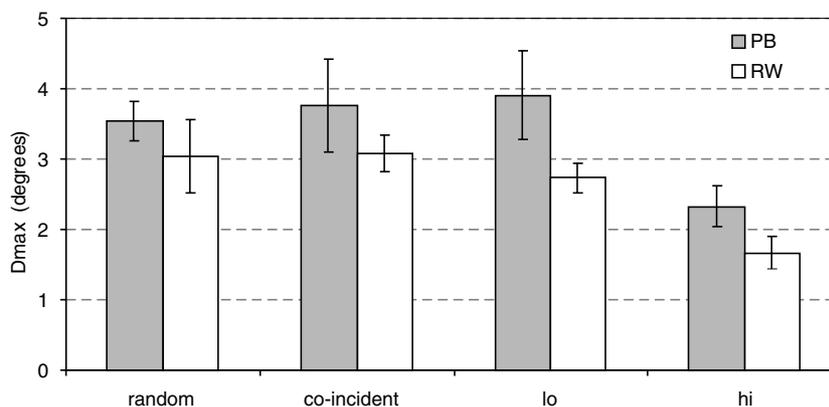


Fig. 3. The upper displacement limit ( $D_{\max}$ ) for filtered natural scenes for two observers (PB and RW). The  $x$ -axis shows the filter condition (see Fig. 2 for details), the  $y$ -axis shows  $D_{\max}$  (in degrees). Random images were the sum of low SFs ( $F_{\text{peak}} = 1$  c/deg) from one image and high SFs ( $F_{\text{peak}} = 6$  c/deg) from another, producing transparent-like images (as in Fig. 2h). Coincident images were the sum of low and high SFs from the same image from another, producing solid images (as in Fig. 2g). Lo and hi images contained only low SFs (as in Fig. 2e) or high SFs (as in Fig. 2f) respectively. Error bars show 95% confidence intervals.

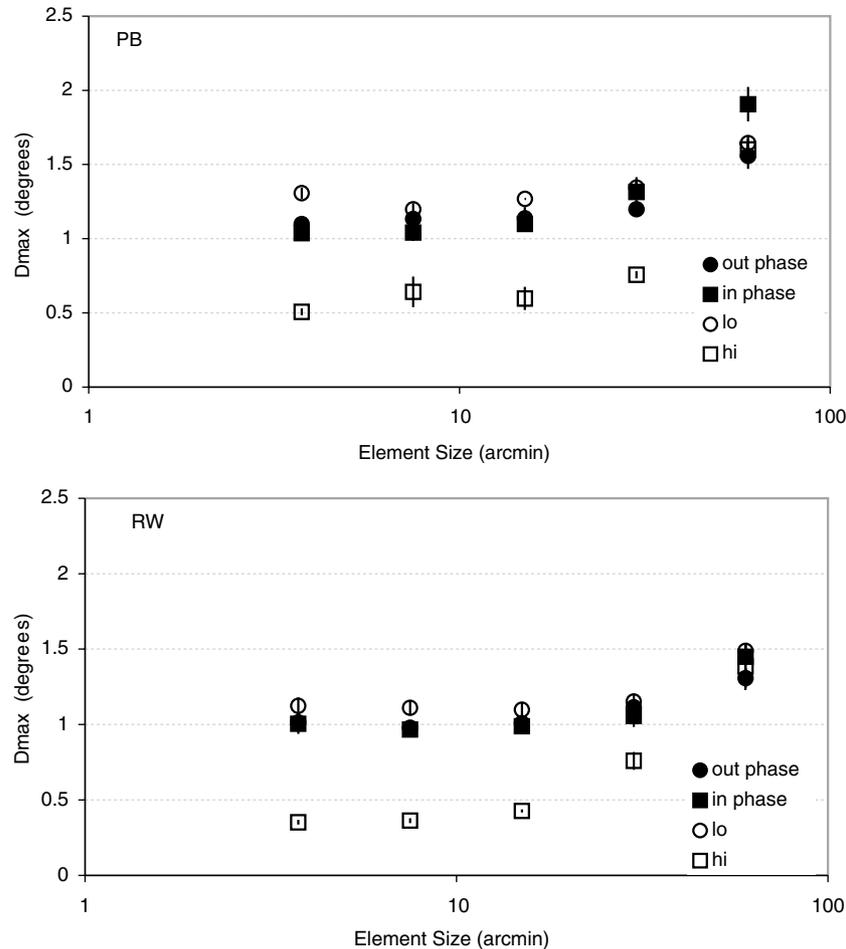


Fig. 4. The upper displacement limit ( $D_{\max}$ ) for filtered binary noise images as a function of element size for two observers (PB and RW). The x-axis shows the size (in arcmin) of the elements forming the binary noise image, the y-axis shows  $D_{\max}$  (in degrees). Open symbols show data for conditions in which images contained a single spatial frequency band: open circles show low SFs ( $F_{\text{peak}} = 1$  c/deg) and open squares show high SFs ( $F_{\text{peak}} = 6$  c/deg). Filled symbols show results for images containing both low and high spatial frequencies; frequencies were either drawn from the same image (filled squares) so that there is a tendency for spatial structure to be coincident (see Fig. 2 column 3), or from independent image samples (filled circles; see Fig. 1 column 4) so that structure at high and low and high spatial frequencies tended to be decorrelated. Error bars show 95% confidence intervals.

$D_{\max}$  was not correspondingly affected by this change in edge frequency as expected from edge-matching principles.

### 2.2.1. Were subjects ignoring high spatial frequencies?

We were concerned that the results were consistent with the strategy of simply attending to low spatial frequency structure in the broad SF images, which can affect object recognition in categorisation studies (Oliva & Schyns, 1997; Schyns & Oliva, 1994). We considered this unlikely because the conditions were randomly interleaved so that observers would have had to attend to high spatial frequencies on some trials and not on others. However, we ran an additional experiment in which we simultaneously collected  $D_{\min}$  (the minimum spatial displacement for correct direction discrimination on 75% trials) and  $D_{\max}$  on randomly interleaved trials. As

optimal sensitivity for  $D_{\max}$  is mediated by low spatial frequencies and  $D_{\min}$  by high spatial frequencies, we would expect performance to degrade for one of these estimates in a mixed task experiment if observers were attending to a single spatial scale. The procedure was as before and employed binary noise images with an element size of  $0.25^\circ$  and two motion conditions ( $D_{\max}$  and  $D_{\min}$ ), combined with the same four filter conditions (low SF, high SF, broad SF coherent, and broad SF incoherent) randomly interleaved per run. In pilot runs we found that  $D_{\min}$  was sometimes less than a single pixel under our original viewing conditions. We therefore increased our resolution by quadrupling the viewing distance to 344 cm and doubling the stimulus size to  $512 \times 512$  pixels, which changed the retinal image size of the stimulus from  $8^\circ$  to  $4^\circ$ . The centre frequency of the band-pass filters remained at 1 and 6 c/deg.

The results are shown in Fig. 5a and confirm the significant difference between conditions (repeated measures ANOVA,  $F_{(3,3)} = 109.8$ ,  $p < 0.01$ ) found with natural images.  $D_{\max}$  is lowest for high SF only noise ( $p < 0.01$ , by linear contrasts) and not significantly different among low SF, broad SF coherent and broad SF incoherent noise. The reduction in the retinal size of the stimuli lowered the absolute value of  $D_{\max}$  and brings our values closer to those of (Morgan, 1992).

Fig. 5b shows the results for  $D_{\min}$ . Again there was a significant difference between conditions (repeated measures ANOVA  $F_{(3,3)} = 14.7$ ,  $p < 0.05$ ) where  $D_{\min}$  is highest for low SF (noise  $p < 0.01$ , by linear contrasts) and not significantly different among low SF, broad SF coherent and broad SF incoherent noise. The naïve observer (RW) was significantly less sensitive to minimum motion tasks than the more experienced observer (PB). These results suggest that the low value of  $D_{\min}$  in the broad SF images, was determined by high SF content in these stimuli and therefore that observers were not simply ignoring high spatial frequencies in our stimuli.

### 3. Experiment 2: Motion sensitivity and across spatial frequencies

The results of Experiment 1 suggest that under our conditions, motion detection in broad-band images is based on the responses of spatial frequency selective motion detectors that operate independent of detectors that are tuned for other spatial frequencies and not on feature-based processes. We reasoned that on this basis motion should not be visible between displaced image pairs whose spatial frequency content does not overlap because different populations of spatio-temporal filters will respond to the two animation frames. In order to test this conclusion, we examined motion perception in RDKs composed of two band-pass filtered images. The source image on each frame was the same, but the peak spatial frequency of the 1-octave band-pass filter was varied between frames so that we could control the spatial frequencies that were common to both animation frames.

Methods were as in Experiment 1 except for the following changes to the stimuli. Random noise images

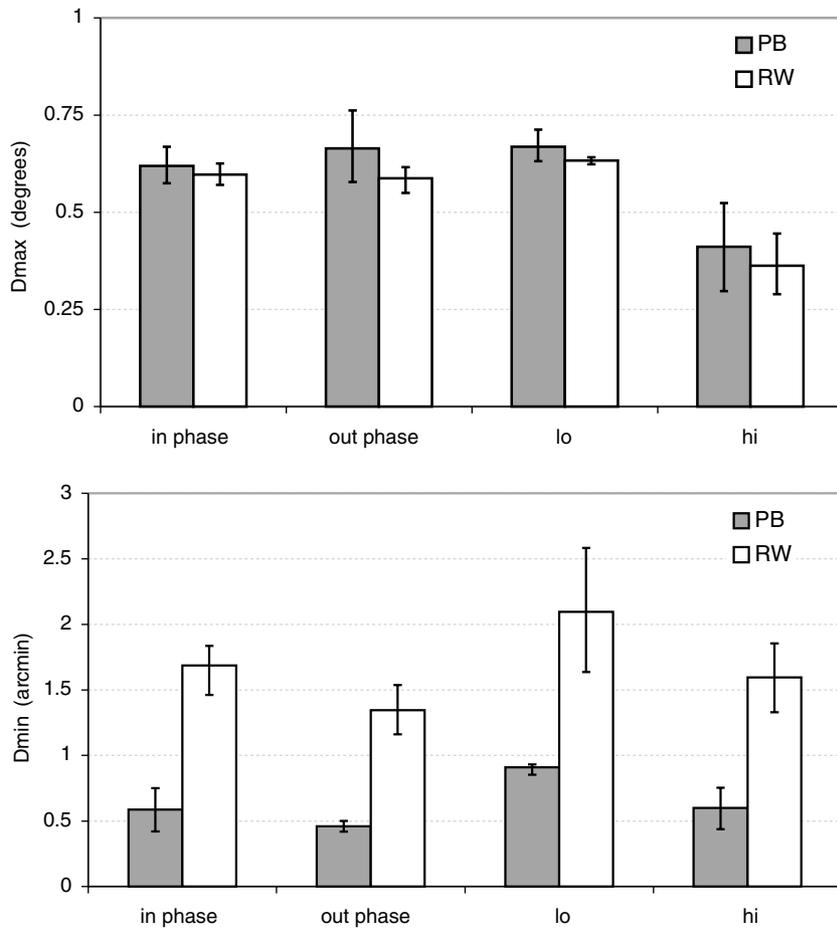


Fig. 5. (a) The upper displacement limit ( $D_{\max}$ ) and (b) lower displacement limit ( $D_{\min}$ ) for filtered noise stimuli for two observers (PB and RW). The x-axis shows the filter condition (see Fig. 2 for details). Error bars show 95% confidence intervals.

were generated as before at three element sizes ( $0.0625^\circ$ ,  $0.25^\circ$  and  $1^\circ$ ), the noise image and a displaced version of it were filtered with log-exponential filters at differing peak frequencies. The peak frequencies of the first and second filters were equally log-spaced around one of four testing frequencies: 0.5, 1, 2 or 4 c/deg. Testing was centred around these four spatial frequencies because a number of previous studies have indicated that motion detection may favour low spatial frequencies under some conditions (Bex & Dakin, 2002; Yang & Blake, 1994). The spacing between the filter peaks was under the control of a QUEST staircase that concentrated observations around the 75% correct point. The displacement of the RDK was fixed at one half the size of an element, which produced 100% correct direction discrimination when there was no spatial frequency difference between RDK frames. Thus observers saw a RDK composed of two frames of a displaced noise image, each with slightly different spatial frequency

content. When the direction of shift was correctly detected, QUEST tended to increase the difference between the peaks of the two filters (while maintaining the same central testing frequency). QUEST tended to decrease the difference following incorrect responses. All other viewing conditions were the same as in Experiment 1, so there were 32 trials per staircase, all 12 conditions were randomly interleaved in a run and the data over a minimum of four runs were combined to calculate thresholds at the 75% correct point.

### 3.1. Results and discussion

Fig. 6 shows the maximum separation between the peak spatial frequencies of two band-pass filtered RDK frames that supports motion at three element sizes and four centre spatial frequencies. Factorial ANOVA showed a significant interaction between these factors ( $F_{(6,6)} = 11.605$ ,  $p < 0.0025$ ). Motion can therefore be

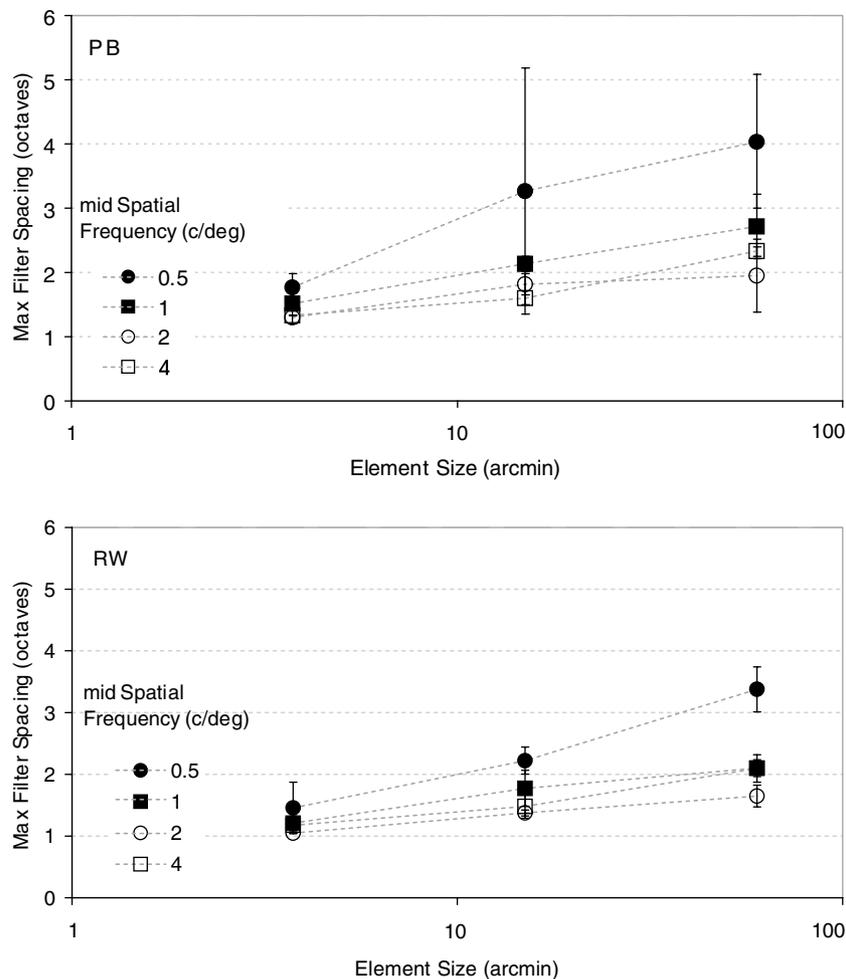


Fig. 6. The maximum tolerable spatial frequency difference between two noise images that support motion perception for two observers (PB and RW). A RDK was shifted by one half the element size (shown on the x-axis). The two frames of the RDK were band-pass filtered at different spatial frequencies that were equally log-spaced about a central value (filled circles: 0.5 c/deg, filled squares: 1.0 c/deg, open circles: 2.0 c/deg and open squares: 4.0 c/deg). The maximum separation between the peaks that supported motion perception is shown on the y-axis (in octaves). Error bars show 95% confidence intervals. See text for details.

detected between band-pass filtered images whose peak spatial frequencies can be separated by as much as 4 octaves. The data show that the separation increases with element size (Fisher's PSLD test showed significant differences between all element sizes,  $p < 0.05$  in all cases), but only the lowest spatial frequency differed significantly from all others ( $p < 0.05$  in all cases). At smaller element sizes, motion can be detected up to a difference of approximately 1 octave. This is in close agreement with a similar study by Ledgeway (1996), the most comparable condition between the studies being with elements subtending  $0.125^\circ$  in the present case, Ledgeway's stimulus subtended  $0.34 \text{ deg}^2$  (at  $\pm 1$  stdev of the 2D Gaussian window) and included 25 dots. The overlapping area of the filters decreases as the separation between the peaks increases (from 35% at 1 octave, 7% at 2 octaves, 0.6% at 3 octaves to 0.01% at 4 octaves). Here we find that observers are able to detect the displacement of dots with as little as 0.01% overlap in spatial frequency when the elements are sufficiently large. Under these conditions, the direction of motion must be detected by a system operating on the location of edges in the stimulus and this mechanism appears not to be selective for the spatial frequencies that define the edges that are integrated.

#### 4. Experiment 3: Motion coherence and structure at high and low spatial frequencies

It is possible that direction discrimination tasks with two flash apparent motion, like  $D_{\max}$ , encourage observers to monitor a single spatial scale, even when the components are weighted for their relative visibility or if the images contain structure at a variety of spatial scales as in natural scenes. Another measure of motion sensitivity, motion coherence (Newsome & Pare, 1988), requires observers to integrate local motion signals over large areas of the retinal image. Recent evidence pertaining to filtered dot stimuli shows that a broad range of spatial frequencies contribute to the global perception of motion, as long as the elements are equated for visibility (Bex & Dakin, 2002). In Experiment 3, we measured motion coherence sensitivity with filtered dot images in which elements defined by low and high spatial frequencies were randomly positioned independently or were coincident or with respect to one another as occurs in real images composed of fewer elements. Signal elements moved in the target direction (leftwards or rightwards), noise elements moved in random directions. A QUEST staircase varied the proportion of signal elements to determine the proportion at which observers could identify the global direction of motion on 75% trials. The data from four runs of 32 trials each were combined to estimate threshold as before. The lifetime of each element was limited to a single displacement, after which it was randomly repositioned in

the display. Each element was initialised with a random starting angle to prevent the flicker that would occur if all elements expired simultaneously. Each element was a band-pass filtered dot with a peak frequency of 1 or 6 c/deg, as in Experiment 1. Each display consisted of 100 elements composed of two groups of 50 elements. The 50 elements in each group were either plotted in the same physical location (coincident), producing 50 tokens on screen; or in random locations, producing 100 tokens on screen. There were four combinations. (1) Low SF: low spatial frequency dots only. (2) High SF: high spatial frequency dots only. (3) Broad SF random: 50 high and 50 low SF dots in random relative locations. (3) Broad SF coincident: 50 low and 50 high SF dots, superimposed. Illustrations of the stimuli are shown in Fig. 2. As in all experiments, stimuli were presented in a central circular window subtending  $8^\circ$  (256 pixels), the edges of which were smoothed with a raised cosine subtending  $0.5^\circ$  (16 pixels). Movies were presented for 506 ms (38 video frames at 75 Hz) and movie frames were updated every 40 ms (three video frames). The onset and offset of the movie was smoothed with a raised cosine temporal envelope lasting 40 ms. A central fixation cross was present throughout.

In order to determine a spatial displacement for each element that did not favour one spatial frequency over another, we measured motion coherence thresholds as a function of displacement size for elements with centre frequencies at 1, 2, 4 and 6 c/deg. Based on these results, shown in Fig. 7, we chose a displacement of  $7.5'$  (4 pixels), which approximately equated motion coherence thresholds for elements with centre frequencies at 1 and 6 c/deg. We also equated the visibility of the elements by equating the RMS contrast of the elements, which is known to approximate contrast detection thresholds (Bex & Makous, 2002) and the supra-threshold apparent contrast of broad-band images (Moulden, Kingdom, & Gatley, 1990) and results in equal contribution to motion coherence (Bex & Dakin, 2002).

##### 4.1. Results

Fig. 7 shows motion coherence thresholds as a function of the spatial frequency and the displacement size of moving dots. The results are in good agreement with a previous study of the effects of low-pass filtering ( $+3.25$  dioptres optical blur) on motion coherence thresholds (Barton, Rizzo, Nawrot, & Simpson, 1996). In such blurred images, thresholds increased for small displacement sizes ( $<16'$ ), but increased at larger displacements ( $>21'$ ), i.e. small displacements are less visible and larger displacements are more visible following blurring. This might be expected from the shift to lower spatial frequencies caused by blurring on the assumption that small displacements are detected by units with fine receptive fields whose input is attenuated by blurring and large

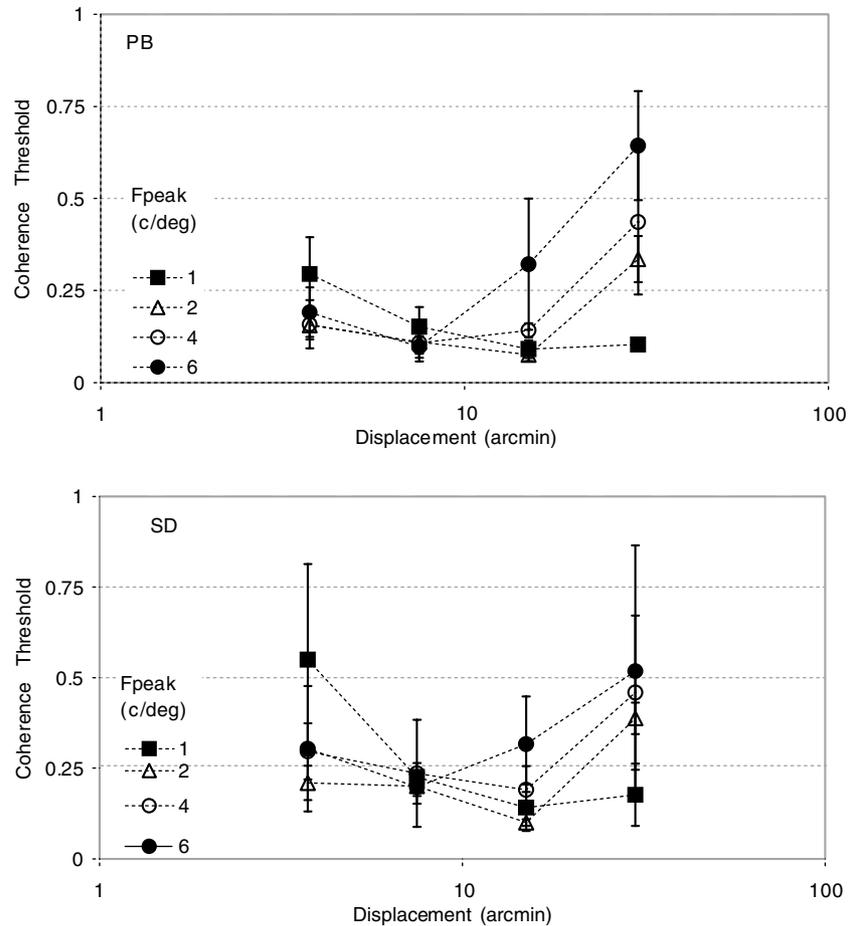


Fig. 7. Motion coherence thresholds, as a function of displacement size and spatial frequency for observers PB and SD. The  $y$ -axis shows the proportion of signal dots required to detect the direction of motion (left or right) of a band-pass filtered noise pattern. The caption shows the four centre frequencies of the band-pass filter in c/deg. Error bars show 95% confidence intervals. See text for details.

displacements are detected by units with coarse receptive fields and blurring attenuates masking from noisy signals at high spatial frequencies (Cleary & Braddick, 1990b). At a displacement of 1/2 cycle of the peak frequency, motion coherence thresholds for band-pass filtered dots are lowest for elements centred at around 3 c/deg (Yang & Blake, 1994). The closest points in our data (3.75'' displacement @ 6 c/deg, 7.5'' @ 4 c/deg; 15'' @ 2 c/deg and 30'' @ 1 c/deg), shows lowest thresholds at 2 c/deg for both observers, in good agreement with this study. For band-pass filtered dots simulating optical flow (forward motion of the observer), directional sensitivity is greatest for filtered elements with a peak spatial frequency of around 1.6 c/deg, when the mean speed of the dots is 2.3 deg/s (Kim & Turano, 1999). This speed corresponds to a displacement of 5.5' displacement size at our 25 Hz movie rate, and for our closest condition (3.8' displacement), we find that a small advantage at 2 c/deg, consistent with their results, allowing for large differences in stimuli (our stimuli contained a single speed, theirs contained a speed gradient) and tasks (left/right judgement versus direction of heading) of the studies.

Motion coherence thresholds for coincident and non-coincident structure are shown in Fig. 8. Open bars show thresholds for coincident elements (50 tokens), filled bars show thresholds for random elements (100 tokens). For both low and high single SF conditions, the threshold number of dots rises when the number of dots increases from 50 to 100, but stay in roughly constant proportion, consistent with previous studies (Baddeley & Tripathy, 1998; Edwards & Badcock, 1994). The similarity in the threshold levels between high and low SF conditions reinforces our selection of displacement and relative contrasts to equate the visibility of the high and low SF groups. Coherence thresholds for mixed high and low SFs that are either coincident or randomly positioned relative to one another are approximately equal.

## 5. General discussion

Convergent evidence from electrophysiological, behavioural and computational studies of visual processing suggests that the first stages of motion processing

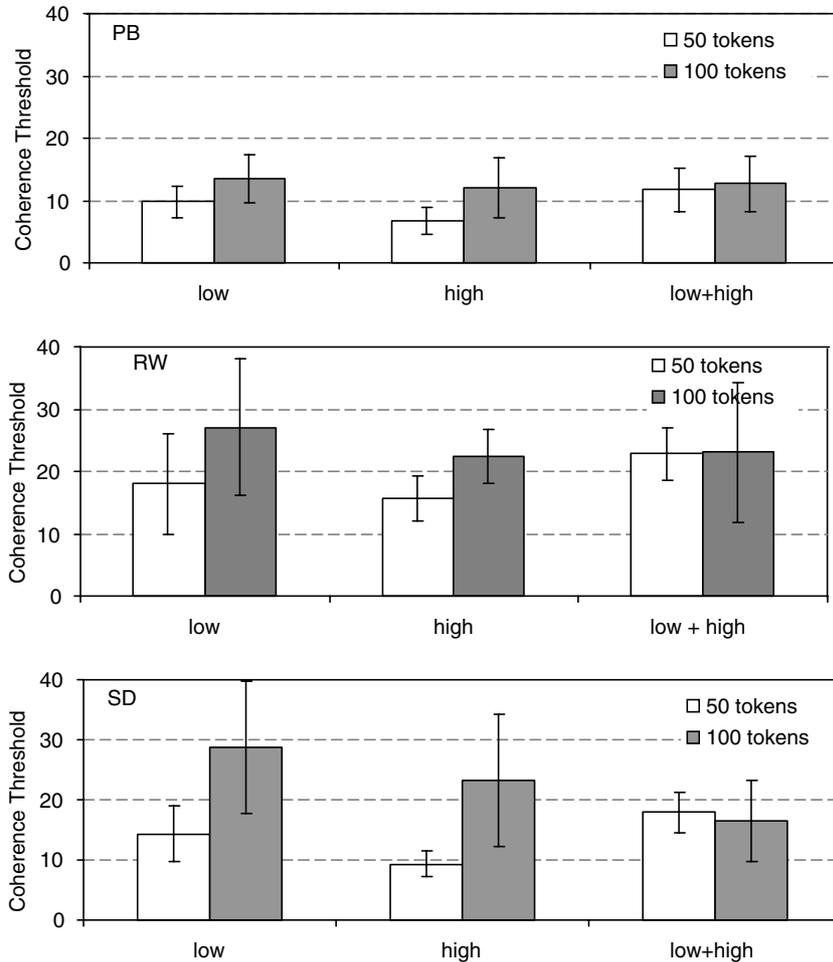


Fig. 8. The effect of spatial coincidence upon motion coherence thresholds for three observers. Threshold was defined as the number of signal dots required to discriminate the direction of motion (left or right) in displays containing 50 (open bars) or 100 (filled bars) “tokens”. Tokens were either composed of a single dot at one centre frequency that was low (1 c/deg; e.g. Fig. 2i) or high (6 c/deg; e.g. Fig. 2j), or were a mixture of low and high spatial frequency dots that were spatially coincident (producing 50 tokens overall; e.g. Fig. 2k) or spatially random (producing 100 tokens overall; e.g. Fig. 2l). Error bars show 95% confidence intervals. See text for details.

involves local analysers that are narrowly selective for the spatial frequency and orientation of objects falling within their receptive fields. In the present studies, we examined how information at different spatial scales is combined in motion perception.

The results of Experiments 1 and 3 show that the relative locations of structures (edges, features, contours, dots) defined by high or low spatial frequencies are unimportant: motion sensitivity seems to be determined by the spatial frequency scale containing the most reliable motion signal. It is possible that the visual system relies on the spatial frequency yielding the lowest estimate of directional bandwidth, in direct analogy with previous studies of texture perception (Dakin & Watt, 1997). This information is at low frequencies for  $D_{\max}$  and high spatial frequencies for  $D_{\min}$ . In motion coherence tasks in which the elements are balanced for visibility, both high and low spatial frequencies appear to contribute equally, but again their relative locations are

unimportant. This is difficult to reconcile with models of motion detection that are based on the correspondence of features in the image (Eagle & Rogers, 1996), possibly following an initial filtering stage (Morgan, 1992; Morgan & Mather, 1994). In order to account for  $D_{\max}$  in high-pass filtered images, the models have to be sensitive to edge structure defined at high spatial frequencies. In our broad-band images where high and low spatial frequency edges are not coincident, the models operate on images with an increased number edges and so, on average  $D_{\max}$  should be lower in the random condition than the coincident condition.

A small number of previous studies have examined interactions across spatial frequencies in motion perception. When two drifting gratings of differing orientation are superimposed, the resultant pattern can appear to slide over one another as two transparent surfaces, or to cohere into a single “plaid” pattern that moves in a compound direction (Adelson & Movshon,

1982). Coherence occurs only when the spatial frequencies of the component gratings are similar (within approximately 1.5–2 octaves), although coherence can occur with greater differences in spatial frequency at lower speeds (Smith, 1992) or when the angle between the gratings is small (Kim & Wilson, 1993), suggesting that interactions across spatial frequencies are relatively weak. However, coherent motion at a low spatial frequency can capture random motion of fine dots or of high spatial frequencies (Ramachandran & Cavanagh, 1987), suggesting that such interactions are possible under some conditions.

Several previous investigations have shown that motion detection is possible in RDKs with differing (but overlapping) spatial frequency content in the two animation frames. When one of the frames is unfiltered and the other is low-pass (Bex et al., 1995; Hess et al., 1998; Morgan & Mather, 1994) or band-pass filtered (Brady et al., 1997), motion energy detection should be possible because of the spatial frequencies common to both images. When the amplitude spectrum of an image is scaled as  $1/F$ , approximately to match the spectra of natural scenes and the sensitivity of the visual system (Field, 1987), motion detection is possible. However, with a binary noise image, whose amplitude spectrum is much flatter, motion can only be detected with modest levels of blur (standard deviations of a Gaussian blurring function of less than around  $20'$ ) (Morgan & Mather, 1994). At least two interpretations could account for these results, either incoherent directional signals at high spatial frequencies mask coherent signals at low frequencies (Cleary & Braddick, 1990b) or the lack of correspondence between edges in the filtered and unfiltered images could cause a correspondence failure (Morgan & Mather, 1994).

When both frames are band-pass filtered, motion detection is possible when the peaks of the band-pass filters are separated by about 1 octave (Ledgeway, 1996). Under comparable conditions with small element sizes, this finding is confirmed in Experiment 2. When the elements are sufficiently large, however, motion detection is possible between pairs of band-pass filtered images whose peaks are separated by as much as 4 octaves. Any overlap in the tails of the band-pass filters under these conditions is extremely small, and so the detection of motion must be based on the correspondence between edges in the two images, even though these edges are defined by very different spatial frequencies.

It may seem paradoxical that Experiments 1 and 3 show that the relative locations of structures defined at different spatial frequencies are unimportant for motion detection, whereas Experiment 2 shows that motion detection can depend on the relative locations of structures defined at high and low spatial frequencies. These apparently contradictory results are best explained by a

unifying approach that encompasses feature-based and energy-based motion detection mechanisms operating under very similar conditions (Smith & Ledgeway, 2001). Spatial frequency selective directional signals can account for the immunity to the relative locations of structure at high and low spatial frequencies, while spatial frequency non-selective edge-tracking operations are required to integrate the motion of edges defined at different spatial scales.

### Acknowledgements

PB was supported by the Wellcome Trust and by BBSRC. SD was supported by the Wellcome Trust and by BBSRC.

### References

- Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2, 284–299.
- Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300, 523–525.
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61, 183–193.
- Baddeley, R. J., & Tripathy, S. P. (1998). Insights into motion perception by observer modelling. *Journal of the Optical Society of America A*, 15, 289–296.
- Baker, C. L. J., & Hess, R. F. (1998). Two mechanisms underlie processing of stochastic motion stimuli. *Vision Research*, 38, 1211–1222.
- Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT press.
- Barton, J. J. S., Rizzo, M., Nawrot, M., & Simpson, T. (1996). Optical blur and the perception of global coherent motion in random dot cinematograms. *Vision Research*, 36, 3051–3059.
- Bex, P. J., & Baker, C. L. (1997). The effects of distractor elements on direction discrimination in random Gabor kinematograms. *Vision Research*, 37, 1761–1767.
- Bex, P. J., Brady, N., Fredericksen, R. E., & Hess, R. F. (1995). Energetic motion detection. *Nature*, 378, 670–672.
- Bex, P. J., & Dakin, S. C. (2002). Comparison of the spatial-frequency selectivity of local and global motion detectors. *Journal of the Optical Society of America A—Optics Image Science and Vision*, 19, 670–677.
- Bex, P. J., & Makous, L. (2002). Spatial frequency, phase and the contrast of natural images. *Journal of the Optical Society of America A*, 19, 1–11.
- Bischof, W. F., & Di Lollo, V. (1990). Perception of directional sampled motion in relation to displacement and spatial frequency: evidence for a unitary motion system. *Vision Research*, 30, 1341–1362.
- Bischof, W. F., & Di Lollo, V. (1991). On the half-cycle displacement limit of sampled directional motion. *Vision Research*, 31, 649–660.
- Boulton, J. C., & Baker, C. L. (1994). Psychophysical evidence for both a “quasi-linear” and a “non-linear” mechanism for the detection of motion. *SPIE*, 2054, 124–133.
- Boulton, J. C., & Baker, C. L., Jr. (1993). Different parameters control motion perception above and below a critical density. *Vision Research*, 33, 1803–1811.

- Braddick, O. J. (1974). A short-range process in apparent motion. *Vision Research*, *14*, 519–527.
- Brady, N., Bex, P. J., & Fredericksen, R. E. (1997). Independent coding across spatial scales in moving fractal images. *Vision Research*, *37*, 1873–1883.
- Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, *26*, 157–170.
- Chang, J. J., & Julesz, B. (1983a). Displacement limits, directional anisotropy and direction versus form discrimination in random-dot cinematograms. *Vision Research*, *23*, 639–646.
- Chang, J. J., & Julesz, B. (1983b). Displacement limits for spatial frequency filtered random dot cinematograms in apparent motion. *Vision Research*, *23*, 1379–1385.
- Cleary, R., & Braddick, O. J. (1990a). Direction discrimination for band-pass filtered random dot kinematograms. *Vision Research*, *30*, 303–316.
- Cleary, R., & Braddick, O. J. (1990b). Masking of low frequency information in short-range apparent motion. *Vision Research*, *30*, 317–327.
- Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*, 3181–3192.
- Eagle, R. A. (1998). Upper displacement limits for spatially broadband patterns containing bandpass noise. *Vision Research*, *38*, 1775–1787.
- Eagle, R. A., & Rogers, B. J. (1996). Motion detection is limited by element density not spatial frequency. *Vision Research*, *36*, 545–558.
- Edwards, M., & Badcock, D. R. (1994). Global motion perception—interaction of the on and off pathways. *Vision Research*, *34*, 2849–2858.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, *4*, 2379–2394.
- Hess, R. F., Bex, P. J., Fredericksen, R. F., & Brady, N. (1998). Is human motion detection subserved by a single or multiple channel mechanism? *Vision Research*, *38*, 259–266.
- Johnston, A., McOwan, P. W., & Benton, C. P. (1999). Robust velocity computation from a biologically motivated model of motion perception. *Proceedings of the Royal Society of London Series B—Biological Sciences*, *266*, 509–518.
- Johnston, A., McOwan, P. W., & Buxton, H. (1992). A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proceedings of the Royal Society of London*, *250*, 297–306.
- Kim, J., & Turano, K. A. (1999). Optimal spatial frequencies for discrimination of motion direction in optic flow patterns. *Vis. Research*, *39*, 3175–3185.
- Kim, J., & Wilson, H. R. (1993). Dependence of plaid motion coherence on component grating directions. *Vision Research*, *33*, 2479–2489.
- Ledgeway, T. (1996). How similar must the Fourier spectra of the frames of a random-dot kinematogram be to support motion perception? *Vision Research*, *36*, 2489–2495.
- Lu, Z., & Sperling, G. (1995). The functional architecture of human visual motion perception. *Vision Research*, *35*, 2697–2722.
- Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London B*, *207*, 187–217.
- Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, *355*, 344–346.
- Morgan, M. J., & Fahle, M. (1992). Effects of pattern element density upon displacement limits for motion detection in random binary luminance patterns. *Proceedings of the Royal Society of London, B*, *248*, 189–198.
- Morgan, M. J., & Mather, G. (1994). Motion discrimination in two-frame sequences with differing spatial frequency content. *Vision Research*, *34*, 197–208.
- Morgan, M. J., Perry, R., & Fahle, M. (1997). The spatial limit for motion detection in noise depends on element size, not on spatial frequency. *Vision Research*, *37*, 729–736.
- Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society of London Series B—Biological Sciences*, *235*, 221–245.
- Moulden, B., Kingdom, F., & Gatley, L. F. (1990). The standard deviation of luminance as a metric for contrast in random-dot images. *Perception*, *19*, 79–101.
- Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *Journal of Neuroscience*, *8*, 2201–2211.
- Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*, 72–107.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research*, *31*, 1337–1350.
- Press, W. H., Teukolsky, A. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C*. Cambridge, England: Cambridge University Press.
- Ramachandran, V. S., & Cavanagh, P. (1987). Motion capture anisotropy. *Vision Research*, *27*, 97–106.
- Reichardt, W. (1961). Autocorrelation, a principle for the evaluation of sensory information by the central nervous system. In W. A. Rosenblith (Ed.), *Sensory communication* (pp. 303–317). New York: Wiley.
- Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges—evidence for time-scale-dependent and spatial-scale-dependent scene recognition. *Psychological Science*, *5*, 195–200.
- Smith, A. T. (1992). Coherence of plaids comprising components of disparate spatial frequencies. *Vision Research*, *32*, 393–397.
- Smith, A. T., & Ledgeway, T. (2001). Motion detection in human vision: a unifying approach based on energy and features. *Proceedings of the Royal Society of London Series B—Biological Sciences*, *268*, 1889–1899.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge: MIT Press.
- van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London B*, *265*, 359–366.
- van Santen, J. P., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, *2*, 300–321.
- Watson, A. B., & Ahumada, A. J., Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, *2*, 322–342.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics*, *33*, 113–120.
- Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research*, *25*, 1661–1674.
- Yang, Y. D., & Blake, R. (1994). Broad tuning for spatial-frequency of neural mechanisms underlying visual-perception of coherent motion. *Nature*, *371*, 793–796.